Managed Care, Physician Incentives, and Norms of Medical Practice:

Racing to the Bottom?

by

David J. Cooper
Department of Economics
Weatherhead School of Management
Case Western Reserve University
10900 Euclid Avenue
Cleveland, OH 44106
djc13@po.cwru.edu

and

James B. Rebitzer
Department of Economics
Weatherhead School of Management
Case Western Reserve University
10900 Euclid Avenue
Cleveland, OH 44106
jbr@po.cwru.edu
National Bureau of Economics Research
Jerome Levy Institute

5/13/02

Abstract

The incentive contracts that managed care organizations write with physicians have generated considerable controversy. If informational asymmetries inhibit patients from directly assessing the quality of care provided by their physician, critics fear that competition will lead to a 'race to the bottom' in which managed care plans use incentive contracts to induce physicians to only offer only minimal levels of care.

We propose a model of managed care competitive strategy to address this issue. The model solves for both physician incentive contracts and product market strategies in an environment of extreme information asymmetry – physicians perceive quality of care perfectly and patients don't perceive it at all. We find that even in this stark setting, managed care organizations are constrained from 'racing to the bottom' by an interaction between physician practice norms (both exogenous and endogenous) and product market competition between plans.

We conclude by considering the implications of our model for public policy designed to limit the influence of HMO incentive systems.

1. Introduction:

Dating back at least to Cournot (1898), students of industrial organization have built formal models to gain insight about product market competition among firms. A similarly rich took-kit of formal models has been developed for understanding incentive systems in organizations.[1] For the most part, these two literatures have evolved independently. Models of product market competition typically ignore the inner workings of firms, while models of incentives in organizations pay scant attention to the firm's competitive environment. This disconnection is unfortunate because, in many settings, the choice of product market and incentive strategies are inextricably linked. If incentives induce the energy and attention required to produce a valuable good or service, then the design of the incentive system should be influenced by the importance of the good or service to the firm's competitive strategy. Likewise, if a firm's cost structure is heavily dependent on the effectiveness of its incentive system, this should influence the choice of a competitive strategy.

In this paper, we consider the interaction between incentive systems and product market competition in a setting of importance for economics and public policy: the delivery of medical services via managed care organizations. By combining a model of a differentiated product oligopoly with a model of physician incentives, we study how managed care organizations shape the style of medical care physicians practice and, as a consequence, the welfare of their members.

---

1 For recent reviews of the incentives literature see Gibbons and Waldman (1999); and Prendergast (1999). For a comprehensive discussion of the literature on physician incentives see McGuire (2000).

When Congress passed the HMO Act of 1973, managed care was considered an alternative delivery system in need of Federal encouragement (Fox, 1997; and Gosfield, 1997). Since that time, managed care has become the dominant form of health insurance in the United States.[2] Firms in this industry adopt a bewildering variety of organizational forms and acronyms such as: health maintenance organizations (HMOs); preferred provider organizations (PPOs); and point-of-service plans (POSs).[3] The common characteristic of these plans and organizations is the combination an insurance function with systems for managing the actions physicians take on behalf of their patients. To help fix ideas, we will focus the subsequent discussion on HMOs, in particular we model the network HMOs that have become an important part of the managed care industry.[4] In contrast to the "staff model," HMOs in which physicians are hired as employees, independent practice HMOs typically do *not* employ physicians. Rather they contract with individual physician practices (or with associations of such practices) to provide medical services. The contracts that independent practice HMOs write are usually non-exclusive, meaning that individual physicians in the network can see patients who are not members of the HMO. The contracts typically also include monetary rewards for successfully controlling medical utilization costs. These cost control incentives are the source of intense public controversy and high stakes litigation (Gosfield, 1997) The core issue in the public debate stems from the information asymmetries that prevent patients from directly evaluating the quality of care provided by their physician. If patients cannot adequately assess care quality, critics fear

---

2 Ma and McGuire (forthcoming) cite evidence that roughly 75% of the privately insured in the U.S. receive health care under some form of management by their health plan.

3 See Fox (1997) and Robinson (1999) for institutional background.

that competition will lead to an HMO led 'race to the bottom' in which physicians operate under severe cost controls and managed care plans offer only minimal levels of care.[5]

To address this issue, we propose a model of the managed care market place that solves for both physician incentive contracts and HMO product market strategies in an environment of extreme information asymmetry—physicians perceive the quality of care they offer perfectly and their patients do not perceive it at all. Even under these stark conditions, we find that competition between HMOs for patients and for physicians will likely constrain 'racing to the bottom'.

At the center of our analysis is the process by which HMOs assemble physician networks and the role these networks play in the competition for customers.[6] Thus we begin by considering how customers go about choosing an HMO. If information asymmetries were *not* a problem, so that potential HMO members could directly assess the style of care offered by a managed care plan, one would expect to see product differentiation. Those willing to pay for a more resource intensive style of care would opt for HMO networks characterized by weak

---

4 These entities comprise one of the largest segments of the managed care market, roughly 40 percent of total HMO enrollment in 1998 (InterStudy, 1999)

5 While our exposition focuses on network HMOs, our key conclusions can extend to any type of managed care organization. For this reason, we drop the term "independent practice HMO" in favor of the more parsimonious term HMO.

6 In the market for commercial health insurance, the "customers" for whom the HMOs compete are typically employers who purchase insurance on behalf of their employees. Competition for employees will compel employers to choose policies with features their employees desire. Indeed, in a perfectly competitive labor market, employers would choose insurance plans with the same combination of cost and quality that the marginal employee would choose for herself. Thus, we simplify our discussion by leaving the employer out of the picture. We treat each individual employee as if they were the HMO's direct customer.

physician incentives and high premiums.[7] Conversely, those who place a relatively greater weight on cost would prefer networks with higher-powered incentives and low premiums. It is not, however, obvious that this logic applies when there is asymmetric information about styles of care offered by HMOs.

In what follows we demonstrate that differentiated product market competition can occur even when HMO members *cannot* observe the clinical attributes of the medical care they receive. We consider a setting in which potential HMO members have preferences over the cost of insurance and the size of the HMO's physician network. Members prefer larger networks because they are more likely to find a physician they like in larger networks. Members differ, however, in their willingness to pay for a larger network.[8] Thus, an essential determinant of competitive strategy for HMOs will be the costs of assembling large physician networks.

In our model, physicians dislike working under high powered, cost-containment incentives in part because these incentives are effective at inducing physicians to adopt lower cost practice styles. If physicians differ in how averse they are to medical cost-containment, then attracting large numbers of physicians to a network will require that the HMO write relatively low-powered incentive contracts. The result, as we will demonstrate below, is differentiated product market competition -- but here differentiation occurs via the size of the physician

---

7 In general, patients desire resource intensive styles of care because additional resources can enhance the expected quality of clinical outcomes. In some cases, however, expensive styles of care may be desirable to the patient even if the likely effect on clinical outcomes is negligible. Extensive use of tests and specialists, for example, may have the effect of reassuring patients they are getting the best care, even if expected clinical outcomes are not improved.

8 Even though members cannot observe the quality of clinical care, they may have non-clinical preferences about the individual physicians they see, e.g. location, age, gender, or communication style.

network rather than direct consumer assessments of the quality of care. Customers who place a relatively high value on physician choice and a relatively low value on the cost of insurance will choose HMOs with large networks and low-powered incentives. Conversely, customers who place a high value on low cost insurance will choose HMOs with small physician networks and high-powered incentives. As we show below, product differentiation by size of network makes an HMO led 'race to the bottom' unlikely, even if patients cannot observe the quality of care they receive. This result is driven by the interaction between incentives and product market competition – consumers demand large physician networks, and HMOs can only provide these networks by weakening cost-containment incentives.

Once we derive equilibrium incentive contracts and medical practice styles, we use our model to examine the efficacy of two regulatory strategies for limiting the influence of physician incentive contracts: (1) capping the proportion of 'at risk' income in physician contracts; and (2) making HMOs legally liable for the adverse medical consequences attributed to their cost-containment systems.[9] We find that both these policies have the intended effect of reducing cost-pressures on physicians, but at the price of increasing the cost of medical services and decreasing the proportion of the population with insurance coverage. Thus, the welfare implications of these policies are ambiguous. Some members, particularly those with a high willingness to pay for physician choice, are made better off by these policy interventions while others, including the newly uninsured, are made worse off.

---

9 The first of these policies was implemented in 1997 (Gosfield, 1997). The impetus for the second policy stems from the 1974 Employee Retirement Income Security Act (ERISA). ERISA provides very little regulatory control over employee health plans and only minor remedies for enrollees. Because ERISA preempts state laws, changes to ERISA have been proposed to

All of our conclusions regarding quality of care depend on physicians having appropriate preferences about the style of medicine they practice. In our model, we refer to these preferences as "norms of medical practice" and treat them as having both an exogenous and endogenous component. [10] The exogenous component of norms specifies the minimum cost practice style a physician will tolerate. This minimum level of care is determined by exogenous factors such as the state of medical knowledge, medical ethics, and the threat of malpractice suits. The endogenous component of norms is based on relative, rather than absolute, levels of care. We posit that physicians are averse to offering less generous care to some patients solely on the basis of these patients' HMO affiliation. Similarly, physicians dislike adopting a practice style that affords less generous care than is provided by physicians elsewhere in the marketplace. Thus, the endogenous component of norms in our model depends not just on physician preferences but also on the effect of the incentive contracts prevailing elsewhere in the market place.[11] We find that the presence of these endogenous norms has important economic implications: they amplify forces that move the market *away* from a 'race to the bottom' while, at the same time, reducing the scope for product differentiation between HMOs.

The plan of the paper is as follows. In the next section, we briefly review the empirical evidence concerning physician incentives. In section 3, we present our model of physician

---

enhance HMO liability for mal-practice (Havighurst, 2000).

10 We use the term "norms" to highlight the fact that physicians derive utility directly from the delivery of appropriate medical care (in addition to the utility derived from the income generated by the provision of medical services). For a discussions of norm based decision making in economic transactions see Kandel and Lazear (1992) and March (1994).

11 The social comparisons in our model of relative practice norms are similar to Encinosa, Gaynor and Rebtizer(2001), Gaynor and Rebitzer's (2001) model of physician medical practices, and to Kranton's model of social identity (Akerlof and Kranton, 2000).

incentives under differentiated HMO competition. Section 4 applies the model to regulatory strategies proposed for the managed care industry. We conclude by discussing directions for future research.

II. Physician Incentives and Managed Care.

In this section, we discuss the economic literature on physician incentives and highlight features of special relevance to the theoretical model presented in Section III.

A number of recent econometric studies suggest that the practice style of physicians is influenced by the explicit and implicit financial incentives under which they operate. Kessler and McClellan (1996), for example, find that reforms in state malpractice laws have an economically and statistically significant effect on patient expenditures for the treatment of heart disease. Specifically, state level initiatives that directly reduced expected damage awards were associated with a 5.3 percent reduction in hospital expenditures on acute myocardial infarctions and a 9.02 percent reduction in hospital expenditures on ischemic heart disease. Since there was virtually no change in clinical outcomes as a result of these cost reductions, the authors interpret their results as evidence of 'defensive' medical practices, i.e. the adoption of tests and procedures whose primary rationale is to reduce liability in the event of a malpractice suit.

Where Kessler and McClellan examined physician responses to changes in the expected cost of malpractice suits, Barro and Beaulieu (2000) examine the response of physicians to changes in the compensation formulas used in their practices. They study the effect of a switch from fixed salary to profit sharing at a set of physician practices owned by a hospital chain. They find that the introduction of a performance-based pay plan increased profitability significantly, primarily because physicians increased the number of patients they saw. In addition changing behavior, Barro and Beaulieu found that the new pay formula altered the composition of physicians practices: the least productive doctors left the company while the new entrants proved to be more productive, on average, than the doctors they replaced.

Barro and Beaulieu's study looked at compensation practices in a fee for service setting. In contrast, Gaynor, Rebitzer and Taylor (2001) examine the effect of cost-containment incentives within an HMO network. The HMO they studied wrote incentive contracts with each of the physicians in their network. These contracts provided substantial monetary rewards to those primary care physicians who kept the medical utilization charges of their patients below target levels. Although the details of the HMO's incentive contracts were complex, the authors were able to identify variation in incentive intensity over time and across individual physicians. The study found that for a typical physician in the network, the physician gained $0.10 in income for every $1.00 reduction in medical utilization costs. This incentive was sufficient to induce a 5 percent reduction in utilization costs.

The common conclusion of these studies, that physician choice of practice style *does* respond to financial incentives, will play a central role in our subsequent analysis.[12] Barro and Beaulieu's additional finding that the terms of incentive contracts can influence a physician's decision to affiliate with a plan or practice will also feature prominently in our model. Finally, Kessler and McClellan's result, that physician's clinical treatment decisions were sensitive to the expected cost of malpractice, will also inform our analysis of the public policies that alter HMO legal liability for malpractice claims.

If physicians are influenced by financial incentives, one would expect to see corresponding effects on the quality of clinical outcomes.   Evidence relating financial incentives to quality of care is, however, scarce and those studies that have directly examined the issue have generally found no effect. In the Kessler and McClellan (1996) study of malpractice reform cited above, changes in tort law that produced dramatic changes in hospital expenditures on heart conditions had no measurable effect on outcome measures.[13] In a later study of heart disease, Cutler, McClellan and Newhouse (2000) compare the treatment for HMO members with others insured by traditional indemnity plans. They find that HMOs have 30% to 40% lower expenditures than traditional plans, but that actual treatments and health outcomes differ little across types of plans. Similar results are found in a study of cost and treatment patterns for Massachusetts state and local government employees. In this paper, Altman, Cutler and Zeckhauser (2000) studied costs and treatment intensity for eight medical conditions.[14] For the 215,000 individuals under 65 years old included in the study, the authors found that average HMO costs were 40 percent lower than those of an indemnity plan offering insurance to the same pool of employees. Across the eight conditions, roughly half of the HMO cost savings were due to the lower incidence of these diseases in the HMO population. Virtually all of the remaining savings come because HMOs pay lower prices for the same treatment. On the basis of this evidence, HMOs do not appear to curb the use of expensive treatments.

The incentives literature indicates that physician practice styles *do* respond to financial incentives, but there is little evidence that HMO cost containment incentives cause a discernable

---

12 Robinson (2001) reaches a similar conclusion based on his review of the medical literature.

13 These outcome measures included 1 year mortality rates, and 1 year readmission rates for either acute myocardial infarction or heart failure.

14 These are heart attacks, certain cancers, diabetes and live births

reduction in care quality -- at least for such serious conditions as heart disease, cancer and diabetes. How can these two, apparently anomalous, findings be reconciled?

One potential explanation (and the one we emphasize in our model) is that norms of professional practice prevent physicians from tolerating incentives that grossly degrade clinical outcomes. An important empirical feature of physician practice norms is that they appear to have a local flavor, i.e. they vary in persistent and meaningful ways from one location to another.[15] These "small area variations" in practice style are something of a mystery to those who study health care because they are not accounted for by variations in underlying clinical conditions, cost of treatment or patient incomes. Some analysts have suggested that these geographic practice patterns are the result of physicians learning by observing the practice style and clinical decisions of other physicians in the vicinity (Phelps, 1992). If so, then practice norms may be at least partially an endogenous result of the contracts prevailing elsewhere in the local market. Put differently, geographic variation in physician styles suggests that norms may be shaped by the competitive environment in which HMOs compete for patients and physicians.

A central premise of our model is that norms of medical practice make high-powered cost-containment incentives distasteful for physicians. Evidence in support of this assumption can be found in a recent survey of physician attitudes published in the <u>New England Journal of Medicine</u>. In this paper the authors state "Our findings suggest that bonuses based on limitation of referrals and on productivity heighten physicians' 'performance anxiety' and their perceptions that care may be compromised in these areas…" (Grumbach et.al. 1998; p. 1520). The same

---

15 For an excellent discussion of this large literature see Phelps (1992).

study also reported that when physicians perceived pressure to limit referrals or improve productivity in ways that compromised care, their satisfaction with their practice declined.

Indirect inference regarding physician preferences with respect to incentives can be found by examining the incentive contracts themselves. If physicians dislike incentives that force them to compromise care, one might expect HMOs to write incentive contracts in ways that mitigate incentive pressure for patients most in need of care. This is exactly what Gaynor et. al. (2001) found in their study of physician incentive contracts in place at an HMO. Specifically, they found that the HMO relied on incentive contracts with built-in safeguards to protect seriously ill patients. For the purposes of calculating cost-containment bonuses, the primary care physicians in the HMO's network were held responsible for all costs incurred by each of their patients, up to $15,000 per year. Should a patient have $20,000 in costs in a year, the physician was held responsible for only the first $15,000. This 'stop-loss' provision was intended to remove cost-containment pressures for seriously ill patients and there is evidence that it had the intended effect. Gaynor et.al. find that the incentive system had no influence on in-patient costs (the largest category of costs for patients in heart disease, cancer etc.). Rather, the cost-containment pressure appears to have had the effect of reducing out-patient expenses and specialist referrals.

In conclusion, the economic literature suggests that HMO incentive contracts do influence physician behavior, but not in ways that degrade care quality for the sickest patients. One explanation for this finding is that HMOs write contracts that take into account physician medical practice norms. In the following section, we present a model of physician incentives in which norms of medical practice constrain an HMO's product market and incentive strategies. The equilibrium described by this model will, in turn, be useful in analyzing the prospects for an HMO led 'race to the bottom'.

III. A Model of Physician Incentives and HMO Competition.

We model competition among HMOs as an extensive form game with three stages. The players in this game are two HMOs and the population of doctors that might treat patients insured by these HMOs.[16] In the first stage of the game, the HMOs simultaneously set the number of doctors they want in their network and the quantity of HMO members they intend to service. Prices for each HMO are then set to clear the market.[17] In the second stage of the game, the HMOs write incentive contracts for the physicians in their network. HMOs are constrained to write contracts that yield the promised number of physicians for their network. In the final stage of the game, doctors make two related decisions: which HMO (or HMOs) to join and what style of medical practice to adopt. Both of these decisions are shaped by the incentive contracts HMOs offer. We use subgame perfection as a solution concept and we solve the model via backward induction. Our exposition of the model, like our solution for the game's equilibrium, begins with the final stage and works backward in time to reach the first stage.

Stage 3: Physician Choice of Network Affiliation and Practice Style: Consider an HMO whose network is composed entirely of primary care physicians. In this HMO, PCPs are "responsible" for the care of their panel of HMO members in both a clinical and economic sense. Clinically, primary care physicians must approve any actions that incur medical utilization costs,

---

[16] We could extend our model to include any number of HMOs as well as old-style indemnity plans, but the key results are easiest to communicate in a model with only two HMOs. To further simplify things we assume that each HMO offers only one plan. Allowing HMOs to offer multiple plans would complicate the model without altering its basic conclusions.

[17] This model of product market competition in the first stage builds off of Gal-Or's (1985) model of differentiated product oligopoly. In the context of HMO competition, the Cournot assumption means that HMOs set target market shares and then set prices to achieve those targets. This focus on market share is roughly consistent with informal discussions about strategy the authors have

e.g. drug prescriptions, referrals to specialists etc. Economically, primary care physicians are also held "responsible", via incentive contracts, for the medical costs incurred by their patients. Managing care by making the primary care physician the "gatekeeper" to resources is common in the managed care industry.

The HMO writes incentive contracts with the primary care physicians in its network. For simplicity we focus on linear contracts having two parameters, a capitation rate and a cost share.[18] The capitation rate for incentive contracts in HMO i, represented by k_i, is a flat fee that the HMO pays the physician for each HMO member in the physician's patient panel. The cost share parameter for HMO i, represented by d_i, is the fraction of incurred medical costs that the physician must bear. Without loss of generality, we assume that it is HMO 1 that will have the low powered incentives and HMO 2 that has the high powered incentives, i.e. that $d_2 > d_1$.

If incentives are to matter, physicians must be free to adopt different styles of medical practice in response to different levels of cost sharing. Think of these medical styles as short-hand descriptions of the strategies primary care physicians use to treat the patients that arrive in their office. For example, a primary care physician may decide to send every case of acne to a dermatologist and every ankle sprain to a sports medicine specialist and every patient with headaches to a neurologist. This style of medical practice would typically generate more medical expenses than one in which the primary care physician tried to treat the acne or the sprains or the headaches themselves. We index practice styles by s and think of s as increasing

---

had with a local HMO. Whatever its face validity, Cournot competition has the virtue of being more analytically tractable than Bertrand competition.

18 In principle the model can be extended to include direct and intrusive monitoring, the informal 'steering' of patients to low-cost providers and more complex contracts, but these features would

with the costliness of a practice style. More specifically we write the cost per patient of a physician adopting practice style s as:

$$c(s) = \beta s^2 \qquad \textbf{(1)}$$

Physicians make choices about their practice style based on a combination of clinical and financial considerations. As professionals, physicians have a sense of what resources patients *themselves* would choose to have spent on their medical treatment if patients had the knowledge and information to make these decisions. In addition, physicians may have intellectual or scientific motives to use the latest and best technology on their patients. Physicians may also practice "defensive medicine" and use tests and procedures to preempt future malpractice suits. Each of these factors will cause physicians to prefer more expensive over less expensive styles. Indeed, if physicians did not generally prefer more expensive practice styles, there would be little need for HMOs to write contracts with incentives for controlling costs.

We combine physician preferences for income with their preferences for more expensive styles of practice and the constraints imposed by the incentive contracts into the following objective function:

$$u_D = \sum_{i=1}^{2} q_i \left[\left(k_i - d_i \beta s_i^2 \right) + \gamma (s_i - \alpha) \right] \text{ with } 0 \le d_i \le 1 \text{; and } \gamma > 0 \ \textbf{(2)}$$

where q_i is the number of patients the physician has from HMO i ; $\left(k_i - d_i \beta s_i^2 \right)$ is the income earned for each member served from HMO i; and $\gamma(s_i - \alpha)$ is the per-member utility derived from the direct returns to adopting a costly practice style.

---

greatly complicate the analysis. See (2001), and Ma and McGuire (forthcoming) for a discussion of other incentive instruments available to HMOs.

In this last expression, α represents the minimum acceptable practice style for the physician, i.e. the doctor would prefer to not join a network rather than using a style that is lower (i.e. less expensive) than α. We stipulate that a physician will join a network if this action meets two criteria. First, being a member of the network must generate non-negative utility when operating with the utility-maximizing practice style (as derived below in (4) – (6)). Second, the utility maximizing practice style must be greater than α, the minimal acceptable practice style.[19] Under our maintained assumption that patients do not have the information or expertise to adequately assess physician practice styles, it is a physician's judgment about what is minimally "acceptable" (or α) rather than the marginal patient's preferences, which limits the physician's choice of styles. We capture heterogeneity in physician judgments by allowing α to be uniformly distributed on the interval 0 to A, where $A > 0$.[20]

We have so far assumed that physicians care only about the style of care they adopt in an HMO. The empirical literature on small area variations in medical practice (touched on in section 2) suggests that physicians are also influenced by the norms of care in their community. We model community standards as emerging from physicians' concern about relative practice norms. We assume that physicians observe the practice style adopted by other physicians in the market and that they experience a reduction in utility when they adopt a practice style that is lower than the most expensive style prevailing in the market. There are at least three reasons

---

19 This formulation has the useful property that all physicians, subject to joining an HMO's network, select the same practice style.

20 For simplicity, we assume that both HMOs use an interior solution – e.g. both HMOs employ incentives that yield a positive number of doctors but do not employ all available doctors. The first statement must trivially hold, since an HMO with zero doctors earns no profits. The second

why physicians may compare their practice style to the maximum style prevailing in the market. First, they may themselves have patients in both HMOs in the market and they may dislike offering less expensive treatment to patients in HMO 2 (where incentive contracts are higher-powered) than HMO 1. Secondly, there may be status-seeking behavior under which physicians seek to offer the latest and most expensive care to their patients. Thirdly, it may be that physicians who practice a style of medicine that is relatively less expensive than other physicians feel more vulnerable to malpractice claims.

Formally, let $\hat{s}$ denote the maximum style present in the local market[21] and $\lambda s_i(\hat{s} - s_i)$ the disutility derived from offering a style below the maximum. The variable λ is positive, implying that so long as $(\hat{s} - 2s_i) > 0$, marginal reductions in practice style reduce physician utility. When $(\hat{s} - 2s_i) = 0$ the disutility of choosing an inexpensive practice style is maximized, so beyond this point we assume the marginal effect of further reductions in s_i is 0.[22] Including the disutility term, we rewrite the utility a physician derives from q_i patients in HMO i as follows:

---

statement implies that the solution for HMO 1's network size, shown in (23), must be less than D. This is a technical requirement with no economic content.

21 In assessing $\hat{s}$, the physician is assumed to *not* include his own style. In other words, $\hat{s}$ is the highest practice style used by *another* physician. While $\hat{s}$ is endogenous to the model, it is exogenous from the point of view of a physician. In terms of verisimilitude, this assumption is harmless, and it allows us to avoid having a discontinuity at $s_i = \hat{s}$. The value of $\hat{s}$ is assumed to be known by all physicians – this is just the standard common knowledge of Nash equilibrium in another guise.

22 We chose this functional form because it allows us to obtain a neat closed form solution for the equilibrium and expresses the key idea that physician utility drops the further away the chosen style is from the maximum. A feature of this functional form is that the marginal disutility of lowering one's style below the maximum decreases the farther one gets from the maximum. In an unpublished appendix, we present an alternative specification that does not have this property. Specifically we model the disutility from practicing relatively low cost medicine as $\lambda(\hat{s} - s_i)^2$ for $\hat{s} < s_i$ and zero otherwise. We derive a closed form solution for this alternative function and

$$u_D^i = q_i(k_i - d\beta s_i^2 + \gamma(s_i - \alpha) - \lambda s_i(\hat{s} - s_i)) \text{ if } s_i > \frac{\hat{s}}{2} \qquad \textbf{(3a)}$$

or

$$u_D^i = q_i\left(k_i - d\beta s_i^2 + \gamma(s_i - \alpha) - \lambda\frac{\hat{s}^2}{4}\right) \text{ if } s_i \le \frac{\hat{s}}{2} \qquad \textbf{(3b)}$$

Having fully specified the utility function, we can now determine the style physicians adopt when choosing to join a HMO. Formally, we are solving for the Nash equilibrium in the proper subgame defined by Stage 3.[23] Conditional on having joined, we get the following first order condition for doctors in HMO i's network.

$$-2d_i\beta s_i + \gamma - \lambda\hat{s} + 2\lambda s_i = 0 \quad i \in \{1,2\} \qquad \textbf{(4)}$$

To close the model, (5) adds in the equilibrium condition that $\hat{s}$ equals the higher of the two practice styles chosen by physicians.

$$\hat{s} = \max[s_1, s_2] \qquad \textbf{(5)}$$

Noting that $s_2 < s_1$ because $d_2 > d_1$ by construction, doctors will choose the following styles in HMO1 and HMO 2:

$$s_2 = \frac{\gamma - \lambda s_1}{2(d_2\beta - \lambda)} \text{ and } s_1 = \frac{\lambda}{2(d_1\beta - \lambda)} \qquad \textbf{(6)}$$

---

show using simulations that it has the same basic properties as the version of the model presented in the body of our paper. We reserve these results for an unpublished appendix because of the very complex equations that characterize the closed-form solution.

23 Because all doctors choose identical styles subject to joining a HMO, we can solve for the equilibrium styles without knowing which doctors join which HMO. Similarly, because utility is linear in quantity of patients served, we don't need to know the equilibrium quantities to determine what styles will be used in equilibrium. Finally, we note that the second order conditions for the physician's maximization problem hold so long as $d_i > \lambda$. Since this is the condition for which the choice of styles is infinite, it will always hold for any incentive contract the HMO would wish to write.

There are two important points to take away from the equations in (6). First, increases in incentives to control costs (represented by increases in d_1 or d_2) have the effect of reducing the costliness of the practice styles physicians adopt. Secondly, the style prevailing in HMO 1 influences the style prevailing in HMO 2. Specifically, if HMO 1 reduces incentives to control costs (and consequently increases s_1), physicians in HMO 2 will also move to adopt more costly practice styles. These spillover effects from endogenous practice norms will turn out to be an important determinant of HMO competitive strategies.[24]

Stage 2: Incentive Contracts and HMO Cost Functions: In this stage of the game, HMOs have committed to attracting a certain number of doctors, δ_i, to their network. Potential HMO members, as we discuss in the next section, value choice in physicians and hence value larger networks over smaller networks. It is natural then to think of the HMOs as being committed to providing a certain number of doctors to their members.

In staff model HMOs, the number of physicians is essentially the number of physicians employed by the organization. In the independent practice HMOs we analyze here, a doctor "joins" a network by agreeing to receive HMO members as patients, thereby accepting the HMO's incentive contract. Specifically, the doctor agrees to receive the capitation rate, k, and pays the cost share parameter, d.

HMO 1 chooses values of k and d to minimize per patient costs subject to the participation constraints that must be met to build a network with δ_1 doctors.[25] Recall that by

---

24 The possibility of spillover effects between HMO networks was first noted and analyzed by Beaulieu (2000).

25 Because costs are linear in the number of patients, minimizing the HMO's cost per patient is equivalent to minimizing costs.

assumption $d_1 < d_2$, implying that physicians in HMO 1's network will employ a higher (or more costly) practice style than those in HMO 2's network. Let D be the total number of available physicians.

$$\min_{k_1, d_1} \; (k_1 + (1 - d_1)\beta s_1^2) \qquad \textbf{(7)}$$

$$\text{subject to } k_1 - d_1\beta s_1^2 + \gamma\left(s_1 - \frac{A\delta_1}{D}\right) \geq 0 \text{ and } s_1 \geq \frac{A\delta_1}{D} \qquad \textbf{(8)}$$

In (8) we place the two physician participation constraints. The first constraint, that physicians must always have non-negative utility under HMO 1's incentive contract, holds with equality. If it did not, the HMO could always reduce costs by cutting k_i.

The second participation constraint states that if HMO 1 wishes to attract δ_1 physicians to the network, it must write a contract such that the optimal style chosen by physicians working under these contracts equals or exceeds what the marginal physician judges to be the minimum acceptable practice style. Since the parameter representing these judgments is assumed to have a uniform distribution between 0 and A, the marginal physician in HMO 1's network of size δ_1 will have $\alpha = (\delta_1/D)A$. Cost minimization requires that this participation constraint must also hold with equality. If it did not, the HMO could always reduce costs further by increasing the share parameter, d_1.[26] Thus it is the marginal physician's judgment of the minimum style a

---

26 To see this, it is sufficient to note that $\partial c/\partial d_1 < 0$ for all $d_1 < 1$. Substituting the first constraint (which holds with equality) from (8) into (7) and differentiating, $\partial c/\partial d_1 = dc/ds_1 * \partial s_1/\partial d_1$. Given that s_1 is a decreasing function of d_1 and costs are an increasing function of s_1, this derivative must be negative.

"reasonable" physician would choose, rather than the marginal HMO member, who determines the style that prevails in the HMO. Substituting (8) into (7) we derive HMO 1's cost function:

$$c_1(q_1,\delta_1) = q_1\beta K\delta_1^2 \text{, where } K \equiv \left(\frac{A}{D}\right)^2. \qquad \textbf{(9)}$$

We derive HMO 2's cost function analogously. HMO 2 chooses d_2 and k_2 to minimize per patient costs while attracting δ_2 physicians into the network.

$$\min_{k_2,d_2}(k_2 + (1 - d_2)\beta s_2^2) \qquad \textbf{(10)}$$

$$\text{subject to } k_2 - d_2\beta s_2^2 + \gamma\left(s_2 - \frac{A\delta_2}{D}\right) + \lambda s_2(s_1 - s_2) \geq 0 \text{ and } s_2 \geq \frac{A\delta_2}{D} \qquad \textbf{(11)}$$

The constraints in (11) are the same as those for HMO 1 except for the final term of the first constraint, $\lambda s_2(s_1 - s_2)$. This term reflects payments that HMO 2 must give physicians to compensate them for the disutility of practicing at less than the highest style. Comparing the participation constraints in (8) and (11), it is clear that every physician who joins HMO 2 will also join HMO 1's physician network, but some physicians in HMO 1's network will refuse to join HMO 2's.[27]

In HMO 2, as in HMO 1, cost minimization ensures that the participation constraints hold with equality. Substituting (11) into (10) we derive the cost function given by (12). Note that HMO 2's costs depend on the size of HMO 1's network. The size of these cost spillovers depends on the strength of physician's relative practice norms: the more potent the relative norms, the greater the spillover

---

[27] It may also be that some physicians do not join either network. We can think of these physicians as working exclusively with patients having traditional indemnity insurance. To simplify the

$$c_2(q_2,\delta_1,\delta_2) = q_2\left(\beta\delta_2^2 + \lambda(\delta_1 - \delta_2)\delta_2\right)K \text{, where } K \equiv \left(\frac{A}{D}\right)^2 \quad \textbf{(12)}$$

Stage 1: Market Equilibrium: We begin by considering the preferences of the consumers in the market. Each consumer can decide to purchase membership in one of the two HMOs or to purchase no health insurance at all. To highlight the information asymmetries that raise fears of a 'race to the bottom', we adopt the strong assumption that *consumers cannot perceive the style of care provided by physicians in the network*. Consumers in our model do, however, have preferences over physicians and anticipate that they are more likely to find a physician they like in larger networks.[28]

For these reasons, we posit that, ceteris paribus, consumers prefer HMOs with large physician networks. More formally, we assign each consumer a parameter, x, that determines the strength of preferences for physician choice and hence network size. This parameter is uniformly distributed over the interval [0, X] with X indicating the greatest preference for access to a larger network of doctors. We write a consumer's utility in terms of premium costs, p, and network size, δ. Since all consumers prefer larger networks ceteris paribus, b is strictly positive.

$$u(\delta,x:p) = bx\delta - p \qquad \textbf{(13)}$$

By identifying the value of x for the marginal member of HMOs 1 and 2, we can use the consumer's utility function to derive inverse demand functions. For convenience, we assume that the total population size is also X. Maintaining our notational convention that HMO 1 has

---

exposition, we do not incorporate this sector into our formal model.

28 Even if patients cannot evaluate the clinical practice styles of physicians, they can still have preferences over physicians because they can evaluate non-clinical attributes of doctors such as location of offices, office hours, age, gender, and communication style. Factors such as a long-

the relatively low -powered incentives, we observe that HMO 1 will also be able to attract the larger network of physicians so that $\delta_1 > \delta_2$. Therefore HMO 2 will set a price such that the marginal member will be indifferent between joining HMO 2 and going without insurance.

$$p_2 = \delta_2 b(X - q_1 - q_2) \quad \textbf{(14)}$$

Likewise, HMO 1 will select a price so that the marginal member of HMO 1 is indifferent between joining either HMO 1 or HMO 2.

$$p_1 = \delta_1 b(X - q_1) - b\delta_2 q_2 \qquad \textbf{(15)}$$

Using these inverse demand functions, we can write the HMOs' profit functions as:

$$\pi_1 = q_1\left(\delta_1 b(X - q_1) - b\delta_2 q_2 - \beta K\delta_1^2\right) \quad \textbf{(16)}$$

$$\pi_2 = q_2\left(\delta_2 b(X - q_1 - q_2) - \beta K\delta_2^2 - \lambda(\delta_1 - \delta_2)\delta_2\right) \qquad \textbf{(17)}$$

The two HMOs simultaneously choose their respective levels of q and δ and each HMO's choice affects the profits of the other HMO. In the Nash equilibrium for this game, each HMO maximizes its profits taking q and δ for the other HMO as fixed. Thus the equilibrium must satisfy the four following first-order conditions:

$$\frac{\partial \pi_1}{\partial q_1} = b(X\delta_1 - 2\delta_1 q_1 - \delta_2 q_2) - K\beta\delta_1^2 = 0 \qquad \textbf{(18)}$$

$$\frac{\partial \pi_1}{\partial \delta_1} = bq_1(X - q_1) - 2\beta K\delta_1 q_1 = 0 \qquad \textbf{(19)}$$

$$\frac{\partial \pi_2}{\partial q_2} = b\delta_2(X - q_1 - 2q_2) - K\beta\delta_2^2 - K\lambda(\delta_1 - \delta_2)\delta_2 = 0 \quad \textbf{(20)}$$

---

term relationship with a particular physician or recommendations from peers may also lead to patients having preferences over physicians.

$$\frac{\partial \pi_2}{\partial \delta_2} = bq_2(X - q_1 - q_2) - 2\beta K\delta_2 q_2 - \lambda K\delta_1 q_2 + 2\lambda K_2\delta_2 q_2 = 0 \qquad \textbf{(21)}$$

It is straight-forward, if tedious, to solve this system of first-order conditions for closed form expressions of each HMO's equilibrium values of q and δ:

$$q_1 = \left(\frac{5\beta^2 - 5\beta\lambda - \lambda^2}{23\beta^2 - 23\beta\lambda - \lambda^2}\right)X \qquad \textbf{(22)}$$

$$\delta_1 = \left(\frac{9b(\beta - \lambda)}{23\beta^2 - 23\beta\lambda - \lambda^2}\right)\frac{X}{K} \qquad \textbf{(23)}$$

$$q_2 = \left(\frac{6\beta^2 - 9\beta\lambda + 3\lambda^2}{23\beta^2 - 23\beta\lambda - \lambda^2}\right)X \qquad \textbf{(24)}$$

$$\delta_2 = \left(\frac{3b(2\beta - \lambda)}{23\beta^2 - 23\beta\lambda - \lambda^2}\right)\frac{X}{K} \qquad \textbf{(25)}$$

Fulfilling the first-order conditions given by (18) through (21) (along with the second-order conditions) guarantees that each HMO's equilibrium strategy is a local maximum of the payoff function given the other's choice. Because the payoff functions are not quasi-concave, this is a necessary but not sufficient condition for Nash equilibrium. In technical notes available from the authors we also prove that equations (22) through (25) represent a global profit maximum for each firm.[29] We also show that the second-order conditions hold.

---

[29] To prove that equations (22) - (25) are a global maximum we must prove that neither HMO can improve its payoff by switching roles, i.e. that (given HMO 1's network) HMO 2 can't do better by building a larger network than HMO 1. Analogously we must also demonstrate that given HMO 2's network, HMO 1 can't do better by building a smaller network than HMO 2. To establish this we derive each HMO's optimal strategy if it switched roles with the other. We then we compare each HMO's maximum payoff after switching roles with the payoff from remaining in the current role. We find that so long as $\lambda \leq 0.5\beta$, the equilibrium in (22) - (25) represents a global profit maximum. Since this also the condition for the HMOs providing differentiated

Combining (23) and (25), we get the following relationship between the equilibrium network sizes of the two HMOs.

$$\delta_2 = \left(\frac{2\beta - \lambda}{3(\beta - \lambda)}\right)\delta_1 \qquad \textbf{(26)}$$

It follows that our assumption of product differentiation ($\delta_2 < \delta_1$), requires that that $\lambda \leq 0.5\beta$. Thus HMOs will engage in product differentiation provided that relative practice norms are not 'too strong' relative to the costs incurred by more generous practice styles.

Having solved for equilibrium values of q and δ for each HMO, it is trivial to solve for equilibrium prices, profits and costs for each HMO. Given that the marginal physician's participation constraints hold with equality for both HMOs, it is similarly easy to derive the equilibrium incentive pay parameters, k and d, and physician practice style, s, for each HMO. Increases in δ *must* be accompanied by a reduction in d and an increase in s.

Discussion of the Market Equilibrium and Comparative Statics: It is immediately obvious from inspection of equations (22) – (25) that the equilibrium strategies in this game do *not* entail a 'race to the bottom' even under extreme information asymmetries in which consumers observe neither physician practice styles nor the incentive contracts HMOs write. HMO 1 charges consumers a relatively high price for the combination of weak physician incentives and large physician networks that yield a more generous style of medicine. HMO 2, in contrast, offers lower premiums and induces its physicians to adopt a more cost-conscious medical style by writing high-powered incentive contracts.

---

products in equilibrium, it holds for all cases of interest.

Our finding that HMOs will not "race to the bottom" is driven by the interaction between HMOs' product market strategies and the incentive contracts they offer physicians. To compete successfully, HMOs must provide their customers with a broad selection of physicians. It follows that HMOs must employ incentive contracts that will induce large numbers of physicians to join their networks. Given physicians' norms of practice, HMOs cannot employ draconian cost-containment incentives and still obtain the needed doctors.[30] Relative practice norms are particularly important in this regard because they raise the prevailing practice style in both HMOs while, at the same time, reducing product differentiation – in effect pushing the market further away from a "race to the bottom."

The strong ties between the competitive environment faced by an HMO and the incentive contract it employs are well illustrated by considering the effect of raising consumers' willingness to pay for larger physician networks, as captured by an increase in b. From equations (23) and (25), we see that increases in b have the effect of increasing the size of both HMO networks and accordingly reducing the intensity of incentives to contain costs.[31] Intuitively, an increase in b raises the pressure on HMOs to provide large networks and hence weakens their ability to use strong cost-containment incentives.

---

30 Our finding that product market differentiation and physician practice norms mitigate a "race to the bottom" is not likely to be undone by adverse selection. In numerical simulations available from the authors, we demonstrate that in settings where x, the parameter governing an individual consumer's willingness to pay for large networks, is positively correlated with the HMO member's expected medical costs, the HMOs still engage in differentiated product market competition in which HMO 1 offers larger physician networks and lower powered physician incentives than HMO 2.

31 Changes in b have no effect on q_1 and q_2, the number of members signing up for each HMO. This result stems from simplifying functional form assumptions and should not be considered a fundamental feature of the economic setting.

To understand the role played by relative practice norms, consider the effect on equilibrium strategies of increasing the potency of physician relative practice norms, i.e. of increasing the parameter λ. Differentiating (22) through (25) with respect to λ and maintaining the necessary and sufficient condition for product differentiation ($\lambda \leq 0.5\beta$) we get the following:

$$\frac{dq_1}{d\lambda} = \frac{-18\lambda\beta(2\beta - \lambda)X}{\left(23\beta^2 - 23\beta\lambda - \lambda^2\right)^2 K} < 0 \quad \textbf{(27)}$$

$$\frac{d\delta_1}{d\lambda} = \frac{9b\lambda(2\beta - \lambda)X}{\left(23\beta^2 - 23\beta\lambda - \lambda^2\right)^2 K^2} > 0 \quad \textbf{(28)}$$

$$\frac{dq_2}{d\lambda} = \frac{-3\beta\left(23\beta^2 - 50\beta\lambda + 26\lambda^2\right)X}{\left(23\beta^2 - 23\beta\lambda - \lambda^2\right)^2 K} < 0 \quad \textbf{(29)}$$

$$\frac{d\delta_2}{d\lambda} = \frac{3b\left(23\beta^2 + 4\beta\lambda - \lambda^2\right)X}{\left(23\beta^2 - 23\beta\lambda - \lambda^2\right)^2 K^2} > 0 \quad \textbf{(30)}$$

From (30), we find increasing the importance of relative practice norms has the effect of causing HMO 2 to increase the size of its physician network. This is to be expected as heightened relative practice norms increases the cost of getting physicians to adopt a practice style less generous than HMO 1's. As HMO 2's network size increases, cost-containment incentives are reduced and premiums rise. HMO 1 responds to HMO 2's incursion into the "up scale" insurance market by increasing the size of its own network as shown by (28). From (27) and (29) we see that an increase in λ reduces the number of consumers choosing either HMO 1 or HMO 2. This implies that some consumers are switching form HMO 2 to the no-insurance option and some are switching from HMO 1 to HMO 2. A final comparative static result can be found by differentiating (26):

$$\frac{d(\delta_2/\delta_1)}{d\lambda} = \frac{\beta}{(\beta - \lambda)^2} > 0 \qquad \textbf{(31)}$$

When relative practice norms increase in importance, product differentiation falls as the low cost HMO becomes more similar to the high cost HMO.[32]

Taken together, the results from equations (27) through (31) demonstrate that relative physician practice norms influence an HMO's incentive contract and product market strategies. Specifically, these norms work against the competitive forces that might otherwise cause HMOs to offer uniformly low levels of care. This conclusion raises a natural question: if these physician norms protect consumers from a 'race to the bottom' do they also increase consumer welfare?

The answer turns out to be ambiguous. Since an increase in λ reduces membership in *both* HMOs, there must occur a corresponding increase in the number of uninsured. These newly uninsured are, by revealed preference, worse off than before the increase in λ. After all, they had the opportunity to opt out of insurance before the increase in premiums and preferred to purchase insurance. However, it can also be shown that utility increases for consumers with a sufficiently strong willingness to pay for physician choice.[33] Other consumers with less willingness to pay for large physician networks are made worse by increasing relative practice norms, even though they remain insured. For these consumers, the increase in premiums

---

[32] We do not place much emphasis on profits in our analysis, because the results are sensitive to functional form assumptions. In the context of these assumptions, however, we can show profits of HMO 1 increase as λ increases. The profits of HMO 2, in contrast, are not a monotonic function of λ.

[33] The welfare implications of increasing λ for those who remain insured are worked out in Appendix 1.

resulting from an increase in λ will lead to a reduction in utility even though they gain access to more physicians. To summarize, an increase in physician practice norms helps consumers who care most about access to large networks and harms those who are most sensitive to prices.

IV. Public Policy.

Concern over the adverse consequences of 'managed care' has grown with the increased importance of HMOs in the U.S. healthcare system. Although the managed care industry has always been subject to regulation at the federal and state level (see Robinson, 1999), there is growing interest in public policy that more directly influences HMO incentive systems (Gosfield, 1997).

In this section, we consider two broad regulatory strategies for shaping physician incentives: (1) imposing caps on the proportion of 'at risk' income allowed in physician contracts; and (2) making HMOs legally liable for the adverse medical consequences attributed to their cost-containment systems. The first strategy is embodied in physician incentive plan (PIP) regulations implemented in 1997 by the Health Care Financing Administration (Gosfield, 1997). These regulations required that incentive contracts could not place more than 25% of physician income 'at risk', i.e. no more than 25% of income could be linked to performance objectives.[34] The second strategy is embodied in proposals to modify the Employee Retirement and Income Security Act (ERISA) in order to make HMOs liable for damages linked to their cost-containment systems (Havighurst, 2000).[35] Some of these proposed changes to ERISA

---

[34] Gaynor, Rebitzer and Taylor (2001) document a case in which these regulations substantially weakened an HMO's incentive contracts. They also present some evidence that these weaker contracts increased the HMOs medical utilization costs.

[35] ERISA offers both limited regulatory control over health plans and only limited remedies. It is

have, in recent years, been included in various proposals for "Patients Bill of Rights" or "Patient Protection Act' legislation (Studdert, Sage, Gresenz and Hensler, 1999).[36]

Analyzing Caps on Physician Incentives: In terms of our model, we can analyze the impact of caps on incentive contracts by considering the effect of a regulation that compels HMO 2 (the high incentive/low cost HMO) to move the incentive parameter, d, below its equilibrium level. We have already demonstrated in Section 3 that cost minimization requires that the two physician supply constraints hold with equality. These constraints imply that HMO 2 will respond to the regulation by altering both it's incentive and product market strategies. On the incentive side, HMO 2 will set d equal to the maximum allowed by law. On the product market side, HMO 2 will increase the size of its physician network to the maximum feasible given the legally mandated cap on d. This response has an intuitive logic. If HMO 2 has to carry the higher medical utilization costs that come with relaxing physician incentives, it will also seek to exploit the chief benefit of lax incentives – the ability to attract more physicians to the network.

Given the one-to-one relationship between incentive intensity and network size, for every incentive cap that binds for HMO 2, there exists a unique, optimal size of physician network. Thus we can treat incentive caps as if they were rules mandating an increase in the size of HMO 2's network to the maximum feasible under the legislatively determined level of d. Formally, let

---

difficult for states to impose tougher regulations because ERISA preempts state law (Havighurst, 2000). Havighurst (2001) discusses other legal strategies that may be used to bring suit against HMOs

36 A third regulatory strategy would be to require that HMOs reveal the details of their physician incentive contracts to patients. Meaningful patient disclosure is limited by the complexity of incentive plans (Gaynor, Rebitzer and Taylor, 2001) and the limited ability of patients to understand even basic information about incentives (Miller and Horowitz, 2000; and Hall, Kid

δ_2^* be the mandated size for HMO 2's network. We assume this constraint is binding – in other words, HMO 2 is forced to use weaker incentives than it would in the absence of the constraint. To get equilibrium values for q_1, δ_1, and q_2, we would like to solve the system of equations given by (18) through (21) but with (21) replaced by the constraint $\delta_2 = \delta_2^*$. Closed form solutions to this problem are too complex to present in this paper, but for policy purposes all we need are derivatives for q_1, δ_1, and q_2 with respect to δ_2^*. These derivatives turn out to be relatively simple:[37]

$$\frac{dq_1}{d\delta_2^*} = \frac{-2\beta K\left(2bq_2 - (\beta - \lambda)K\delta_2^*\right)}{b\left(2b\left(2X - 3q_1\right) - \delta_2^* K\left(2\beta - \lambda\right)\right)} < 0 \quad \textbf{(32)}$$

$$\frac{d\delta_1}{d\delta_2^*} = \frac{2bq_2 - (\beta - \lambda)K\delta_2^*}{2b\left(2X - 3q_1\right) - \delta_2^* K\left(2\beta - \lambda\right)} > 0 \quad \textbf{(33)}$$

$$\frac{dq_2}{d\delta_2^*} = \frac{-K\left((\beta - \lambda)\left(2X - 3q_1\right) - \left(2\beta - \lambda\right)q_2\right)}{2b\left(2X - 3q_1\right) - \delta_2^* K\left(2\beta - \lambda\right)} < 0 \quad \textbf{(34)}$$

Intuitively, the positive relationship between δ_1 and δ_2^* occurs because an increase in the size of HMO 2's physician network reduces HMO 1's advantage with patients who highly value large networks. To regain its comparative advantage, HMO 1 must broaden its own network. For both HMOs, increases in their network size lead to increases in the marginal cost of serving an additional patient. Thus, imposing a binding regulation on HMO 2 has the effect of reducing the total number of members who join any HMO.

---

and Dugan, 2000)

37 In signing the derivatives, we maintain the assumption that we are in the region that generates strict product differentiation in the absence of any constraint; e.g. $\lambda \leq 0.5\beta$.

The welfare effects of capping incentives are similar to those described in Section 3 for increases in the strength of relative practice norms. Since both q_1 and q_2 are decreasing functions of δ_2^*, the number of uninsured individuals must increase as caps on physician incentives become more binding. Those who become uninsured as a result of incentive caps are, by revealed preference, made worse off by the regulations. Of the remaining insured, those with sufficiently high willingness to pay for physician choice will be made better off while others, with a smaller willingness to pay for choice, will experience a decline in utility. Even though these individuals benefit from access to larger networks, this gain is offset by having to pay increased premiums.

Changes In HMO's Legal Liability: We have so far focused on interventions that cap the intensity of physician incentives. We now consider an alternative regulatory strategy: making HMOs liable for medical malpractice linked to their cost-containment systems.

Lawyers use the term 'judgment proof' to describe defendants who lack the resources to pay damages in tort cases. For the most part, physicians are far from judgment proof. Rather they are highly paid professionals who typically carry substantial amounts of medical malpractice insurance. This is why we argued above that the threat of malpractice suits is one of the factors shaping physicians' absolute and relative practice norms. Because HMOs must take these norms into account in order to attract physicians to their networks, the HMOs already bear some of the costs of malpractice. Similarly, some of the expected costs of malpractice are borne by the HMOs' members in the form of higher premiums. It follows from this that simply altering ERISA and the tort system in order to make HMOs directly liable for damages in malpractice suits will, by itself, have *no* effect on the practice styles that characterize HMO networks. Intuitively, the HMO always makes the marginal physician indifferent between joining its network or not. Shifting costs onto the HMO largely results in a decrease in the capitation rate.

ERISA reforms *will* have an effect, however, if the naming of HMOs as defendants has an influence on the magnitude of damages juries award. A recent series of papers examining jury behavior have found that punitive damage awards are influenced not only by the facts of the case but also by the identity of the defendants. Large organizations with deep pockets are typically hit with higher punitive damage awards than smaller organizations (Kahneman, Schkade and Sunstein 1998). If, by virtue of their size and deep pockets, HMO defendants in medical malpractice suits receive larger punitive damage awards than physician defendants do, we can expect changes in ERISA to affect the behavior of HMOs. In the following analysis we discuss two channels through which HMO malpractice liability can influence equilibrium outcomes: (1) if increased jury awards increase fixed costs per patient; and (2) if increased jury awards raise the incidence of various forms of 'defensive' medical practices. As we will discuss below, 'defensive medicine' can be represented in our model in three different ways. In the interest of brevity, we state the likely effect of these changes without providing formal proof (see Table 1 for a summary). To the extent that these results don't follow directly from what has been shown above, proofs are available from the authors upon request.

Increases in Fixed Cost per Patient: By assumption, we have set an HMO's fixed cost of taking on an additional patient equal to zero. We can expect, however, that changes in ERISA leading to higher malpractice damages may appear as an increase in the fixed cost of taking on a patient.

Technically, fixed costs appear as a constant on the right hand side of (18) and (20), the first order conditions for HMO 1 and 2 with respect to quantity. It can be shown that an increase in this fixed cost increases network size and decreases quantity of patients served for HMO 2. The intuition behind this result is straight forward – because quantity has become relatively more expensive, the HMO substitutes from quantity towards larger network size. For HMO 1 the results are reversed, with quantity rising and network size falling. This may seem surprising, but note that the direct effect of increasing fixed costs per patient are smaller for HMO 1 than for HMO 2 since the former serves a relatively small number of patients. This cost advantage for HMO 1 is sufficiently large to reverse the direct effects due to an increase in fixed costs. Although the number of members of HMO 1 increases, the *overall* effect of the increase in fixed costs is a decrease in the total number of patients insured by HMOs. This follows because the decrease in quantity for HMO 2 is larger than the increase for HMO 1. As we discussed above, these newly uninsured are *always* made worse off by the change. The welfare of those remaining insured is ambiguous.

Increase in the Incidence of Defensive Medicine: HMOs may respond to heightened liability for malpractice by incenting their physicians to practice 'defensive medicine', i.e. to spend resources on medical care that has little clinical value, but helps limit liability in the event of a law suit. In terms of our model, an increase in defensive medical practices can be represented in three ways: by a decrease in the effect that costly practice styles have on costs (β); by an increase in the minimum acceptable practice style, α; and/or by an increase in the cost of practicing a relatively inexpensive style of practice, λ. We discuss each of these in turn.

We have already indicated that part of the relationship between choice of practice style, s, and medical costs is due to the expected costs of malpractice suits. If the expected awards in malpractice suits increase, then costly defensive medical practices can actually reduce the marginal cost of adopting these expensive practice styles. In terms of our model, this means that

β falls as expected jury awards increase. This parameter change causes the equilibrium quantity of patients insured to (weakly) decrease for both HMOs and network size to increase for both HMOs. Intuitively, the decrease in β makes expensive practice styles more attractive for HMOs. This effect is sufficiently strong to overwhelm the direct effect on quantity from decreasing costs per patient (due to decreasing β).

A second plausible response of HMOs to changes in ERISA that increase the size of malpractice awards is to increase the generosity of the practice style of physicians in the network. To induce more generous practice styles, HMOs need to weaken the incentive intensity of their physician contracts with resulting effects on network sizes and quantities of patients served. Formally, this change is equivalent to a cap on incentive parameter *d* in the low-cost HMO. We know from the preceding section, what the effects of such a change will be: both HMOs in the market will adopt lower-powered incentives and also build larger physician networks. As a result, each HMO will attract fewer members and the number of uninsured will rise.

The third channel by which changes in ERISA would influence HMO management is via the *relative* generosity of the practice style prevailing in the network. Consider a malpractice case in which a doctor is charged with malpractice for treating a patient in HMO 2 of our model. Imagine that the question at issue is that the physician did not recommend a controversial and expensive procedure that physicians in HMO 2's network do not offer but which physicians in HMO 1's network routinely offer. If the jury finds in favor of the plaintiff, Kahneman et. al's results suggest that the damage awards will be greater if the HMO and the physician are plaintiffs than if the physician alone were the plaintiff. From this motivating example, it is not hard to imagine that changes in ERISA can be especially hard on the low-cost HMO in a market.

By increasing the costs of using a practice style below that of its competitor, this policy change puts pressure on it to weaken its cost-control incentives. We can capture this effect in our model as an increase in λ, the cost of adopting a *relatively* inexpensive practice style. From the analysis in section 3.4, we know what the effects of an increase in λ will be. The low-cost HMO will reduce the incentive intensity of its physician contracts and increase the size of its physician network. HMO 1 will respond to HMO 2's entry into its market segment by doing the same thing: reducing incentives and increasing network size. The net result is that membership in both HMOs fall while the number of uninsured increase.

To sum up, each of the mechanisms through which one might introduce defensive medicine into our model (decreases in β; and increases in α and λ) have different effects on different variables, but they all share one feature: they increase the number of uninsured. We know by revealed preference, that individuals who become uninsured must be made worse off. For individuals who remain insured, those with high willingness to pay for access to large physician networks will benefit while those who are more price sensitive will be harmed.

Discussion of Policy Results: We have considered the likely effects of two broad strategies for limiting an HMO's ability to impose cost containment incentives on physicians. Our results indicate that both of these strategies will have the likely effect of increasing medical costs and increasing the number of uninsured. The newly uninsured are always made worse off by these interventions. The welfare effects on those remaining insured are murky – with some consumers being made worse off and others better off.

Should we conclude from this that regulating physician incentive systems is a bad policy idea? Our answer to this question is no. There are important externalities to health care that may not be fully captured in the factors that constrain HMO behavior: physician practice norms, customer preferences for physician choice, and expected awards for malpractice.[38] Some of these externalities involve health outcomes – if more expensive treatments are also more effective then financing these treatments may improve the welfare of care givers and family members who are not directly involved in the purchase of the health care insurance. Other externalities involve the physician-patient relationship -- the possibility of high-powered financial incentives anywhere in the health system can undermine a patient's willingness to trust in and listen to a doctor's advice. If the net social value of more expensive practice styles or restrictions on contracting exceeds their private value, there remains a strong case for interventions limiting the ways that HMO's regulate care.

The lesson of our analysis is *not* that policy interventions are necessarily a bad idea, but rather that they must be undertaken with an understanding of their cost – especially the increase in the number of uninsured.[39] Policies that regulate HMO incentive systems can be made more effective if they are implemented in conjunction with policies that increase access to care for the uninsured.

V. Conclusion.

In this paper, we have studied the interaction between the incentive contracts that Health Maintenance Organizations (HMOs) write with physicians and product market competition among HMOs. The use of incentive contracts with strong cost-containment provisions is

---

[38] In principle, jury awards in malpractice cases can be a mechanism for introducing these externalities in the decision calculus of the HMO and physicians. In practice, however, research finds that jury awards are likely to be both highly variable and arbitrary (Kahneman, Schkade and Sunstein, 1998).

[39] Ultimately, the magnitude of the increase in uninsured is an empirical question that cannot be assessed without calibrating our model to real world parameters. Studdart et. al. (1999) suggest

controversial because of the information asymmetries that prevent patients from directly assessing the quality of care provided by their physician. The central economic and policy question is whether, in an environment where consumers cannot assess the quality of care they receive, competition among HMOs will lead to a 'race to the bottom' in terms of care quality.

To address this question we develop a model that solves for both physician incentive contracts and managed care product market strategies in an environment of extreme information asymmetry – physicians perceive care quality perfectly and patients don't perceive it at all. We find that even in this stark setting, HMOs are constrained from 'racing to the bottom' by an interaction between physician practice norms (both exogenous and endogenous); and product market competition.

We closed our analysis by considering the implications of our model for public policy designed to limit the intensity of HMO incentive systems. We analyze two such policies: caps on "at risk" physician income and legal changes that make HMOs directly liable for damages caused by their incentive systems. We find that these policy interventions will likely have the effect of weakening HMO incentive systems throughout the market - even for HMOs to which the policies do not directly apply. As a result, medical costs and premia will rise while the number of uninsured will increase. The newly uninsured are clearly made worse off by these interventions even though the welfare effect on those remaining in the insurance system are complex (with some consumers gaining and others losing). It follows from this that policies aimed at regulating HMO incentive systems could be made more efficacious if they also involved actions to improve access to health care for the uninsured.

---

the increase in costs from increased HMO liability is quite uncertain.

Our model highlights a number of still unresolved empirical and theoretical issues that are worthy of further study. Of these, perhaps the most important is understanding the determinants and scope of physician practice norms. At present these norms are only dimly understood by economists. To what extent are norms shaped by such exogenous factors as the malpractice system and the state of medical knowledge and to what extent by such endogenous factors as the practices adopted by others in a physician's reference group? Further research into the formation and operation of norms is critical for understanding the impact of incentives in healthcare and in the many other settings where critical outcomes hinge on the decisions and actions of highly skilled professionals.

Table 1:

Sign of Predicted Effects of Making HMOs Liable in Malpractice Suits

"+" Positive Predicted Effect "-" Negative Predicted Effect

Possible Effect of Shifting Liability to HMOs	Fixed Cost per Patient Increases	Increase in the incidence of 'defensive medicine' due to higher malpractice awards.		
Representation of Policy Change in Formal Model	Fixed Cost↑	$\beta\downarrow$	$\delta_2^*\uparrow$	$\lambda\uparrow$
HMO 1 quantity (q_1)	-	-	+	-
HMO 1 Network Size (d_1)	+	+	-	+
HMO 2 quantity (q_2)	-	-	-	-
HMO 2 Network Size (d_2)	+	+	+	+
Number of Uninsured (q_1+q_2)	-	-	-	-

References:

1.) Akerlof, George A., and Kranton, Rachel E., (2000) "Economics and Identity" Quarterly Journal of Economics. 115:3. p.715-53.

2.) Altman, Daniel., Cutler, David M. and Zeckhauser, Richard J.,(2000) "Enrollee Mix, Treatment Intensity, and Cost in Competing Indemnity and HMO Plans" NBER Working Paper. No.7832. August 2000.

3.) Barro, Jason R and Beaulieu, Nancy Dean,(2000) "Selection and Improvement: Physician Responses to Financial Incentives" No.Harvard Business School Working Paper. July 2000.

4.) Beaulieu, Nancy Dean,(2000) "Externalities in Overlapping Supplier Networks" Harvard University Working Paper. November 2000.

5.) Cutler, David M., McClellan, Mark., and Newhouse, Joseph P., (2000) "How Does Managed Care Do It?" Rand Journal of Economics. 31:3. p.526-548.

6.) Encinosa, W., Gaynor, M. and Rebitzer, J.,(2001) "The Sociology of Groups and the Economics of Incentives: Theory and Evidence On Incentives in Medical Groups" Working Paper Case Western Reserve University. August 2001.

7.) Fox, Peter D., "An Overview of Managed Care" Essentials of Managed Health Care. edited by Peter R. Konstvedt. Aspen Publishers, Inc. :Gatihersburg, Maryland.

8.) Gal_Or, Esther, (1999) "Quality Competition Versus Uninsurance in Health Care Markets." Advances in Applied Microeconomics. edited by Michael Baye. JAI Press Inc.

9.) Gal-Or, Esther. (1985) "Differentiated Industries Without Entry Barriers" Journal of Economic Theory Vol. 37 p. 310-339.

10.) Gaynor, Martin. Rebitzer, James B., and Taylor, Lowell J.,(2001) "Incentives in HMOs" National Bureau of Economic Research Working Paper. No.85222. October 1, 2001.

11.) Gibbons, Robert and Waldman, Michael, (1999) "Careers in Organizations: Theory and Evidence" Handbook of Labor Economics. Volume3B. edited by Orley Ashenfelter and David Card. Elsevier Science, North-Holland :Amsterdam; New York and Oxford:. p.2373-2437.

12.) Gleid, Sherry, (2000) "Managed Care" Handbook of Health Economics. Volume1A. edited by A.J. Culyer and Joseph P. Newhouse. Elsevier :Amsterdam.

13.) Gosfield, Alice G, (1997) "Who is Holding Whom Accountable for Quality?" Health Affairs. 16:3. p.26-40.

14.) Grumbach, K; Osmond, D; Vranizan, K; Jaffe, D; Bindman, A, (1998) "Primary Care Physicians' Experience of Financial Incentives in Managed-Care Systems." New England Journal of Medicine. 339:21. p.1516-21.

15.) Hall, Mark A., Kidd, Kristin E., and Dugan, Elizabeth, (2000) "Disclosure of Physician Incentives: Do Practices Satisfy Purposes?" Health Affairs. 19:4. p.156-64.

16.) Havighurst, Clark C., (2000) "American Health Care and The Law - We Need to Talk!" Health Affairs. 19:4. p.84-106.

17.) Havighurst, Clark C., (July/August 2001) "Consumers Versus Managed Care: The New Class Actions" Health Affairs.

18.) InterStudy,(1999) "The InterStudy Competitive Edge, Part II: The HMO Industry" Interstudy Publications. No.Report 9.1. 1999.

19.) Kahneman Daniel, Schkade David and Sunstein Cass R., (1998) "Shared Outrage and Erratic Awards: The Psychology of Punitive Damages" Journal of Risk and Uncertainty. p.1-53.

20.) Kandel, Eugene and Lazear, Edward P., (1992) "Peer Pressure and Partnerships" Jo urnal of Political Economy. 100:4. p.801-817.

21.) Kessler, Daniel; McClellan, Mark, (1996) "Do Doctors Practice Defensive Medicine?" Quarterly-Journal-of-Economics. 111:2. p.353-90.

22.) Ma, Albert Ching-to. and McGuire, Thomas G., (forthcoming) "Network Incentives in Managed Care" Journal of Economics and Management Strategy.

23.) March, James G., (1994) A Primer on Decision Making: How Decisions Happen. The Free Press : New York.

24.) McGuire, Thomas G., (2000) "Physician Agency" Handbook of Health Economics. Volume1A. edited by Anthony J.Culyer and Joseph P. Newhouse. Elsevier :Amsterdam.

25.) Miller, Tracy E. and Horowitz, Carol R., (2000) "Disclosing Doctors' Incentive: Will Consumers Understand And Value the Information?" Health Affairs. 19:4. p.149-55.

26.) Phelps, Charles E., (1992) "Diffusion of Information in Medical Care" Journal of Economic Perspectives. 6:3. p.23-42.

27.) Prendergast Canice, (1999) "The Provision of Incentives in Firms" Journal Of Economic Literature. 37:1. p.1-63.

28.) Robinson, J C, (2001) "Theory and Practice in the Design of Physician Payment Incentives." Milbank Quarterly. 79:2. p.149-77.

29.) Robinson, James C., (1999) The Corporate Practice of Medicine: Competition and Innovation in Health Care. University of California Press : Berkeley and London.

30.) Studdert, David M., Sage, William . Gresenz, Carole Roan; and Hensler, Deborah R., (1999) "Expanded managed Care Liability: What Impact on Employer Coverage?" Health Affairs. 18:6 p. 7-27

Appendix 1: Endogenous Norms and Consumer Welfare:

We consider the effect of an increase in spillovers on consumer's welfare. Differentiating the consumer's utility function (13) we get the following:

$$\frac{du}{d\lambda} = bx\left(\frac{d\delta}{d\lambda}\right) - \left(\frac{dp}{d\lambda}\right) \qquad \textbf{(A1)}$$

We have already shown that $d\delta/d\lambda > 0$ for members of either HMO and that $dp/d\lambda$ is independent of *x*. From this it follows that $du/d\lambda$ is an increasing function of x. We can prove that $du/d\lambda > 0$ for a segment of the market by identifying the consumer with the minimum value of *x* for which $du/d\lambda > 0$.[40]

To determine the welfare effects of increasing λ for consumers originally in HMO 1, we evaluate $du/d\lambda$ for the consumer in HMO 1 with the weakest willingness to pay for physician choice. This marginal consumer has $x = X - q_1$ and it is easy to show the following:[41]

$$\left.\frac{du}{d\lambda}\right|_{x=X-q_1} = \frac{9b^2X^2\left(26\beta^3\lambda - 39\beta^2\lambda^2 + 15\lambda\beta^3 - \lambda^4\right)}{\left(23\beta^2 - 23\beta\lambda - \lambda^2\right)^3 K^2} > 0 \qquad \textbf{(A2)}$$

Therefore, an increase in relative practice norms benefits all consumers with a willingness to pay strong enough to choose HMO 1.

---

40 On the margin, the worst off consumer in a HMO either switches HMOs or moves to being uninsured as λ increases. For our calculations, we treat this consumer as if she stayed with the same HMO. If this consumer would have been better off staying with the same HMO than she was prior to the change in λ (although not as well off as by switching), all consumers in the same HMO with higher values of x must also be made better off by the increase in λ.

41 This derivative is positive if we maintain the necessary and sufficient condition for product differentiation ($0 < \lambda < 0.5\beta$).

The story is more complex for consumers that initially choose HMO 2. We know that some consumers in HMO 2, those with relatively low values of x, are made worse off by increased λ ($du/d\lambda < 0$) because some of them choose to become uninsured. Similarly, we know from (A2) that some consumers in HMO 2 with relatively high values of x are made better off by increased λ ($du/d\lambda>0$).[42] Since $du/d\lambda$ is a monotonic function of x, there must exist some value of x such that $du/d\lambda|_x = 0$.

Rather than solving for this breakpoint directly, it is simpler to solve for the proportion of consumers using HMO 2 who are made worse off by an increase in λ. Define the function $\hat{x}(x)$ as follows: $\hat{x}(x) \equiv x - (X - q_1 - q_2)$; $\hat{x}(x)$ gives the number of patients in HMO 2 whose willingness to pay for larger networks is less than x. Combining this definition with the utility function given by (13) and the price function for HMO 2 given by (14), we can rewrite the utility function as $u(\delta_2, \hat{x}) = b\delta_2 \hat{x}$ for any consumer in HMO 2. Differentiating u by λ we learn that $du/d\lambda = 0$ when:

$$\hat{x} = \left(\frac{\frac{d(q_1 + q_2)}{d\lambda}}{\frac{d\delta_2}{d\lambda}} \right) \delta_2 \quad \textbf{(A3)}$$

Dividing $\hat{x}$ by q_2, we can solve for the proportion of consumers in HMO 2 who are harmed by an increase in relative practice norms:

---

42 This follows because some consumers in HMO 1 switch to HMO 2 when λ increases, even though their utility would increase if they stayed in HMO 1. It follows that HMO 2 consumers with values of *x* close to those who switch from HMO 1 to HMO 2 will also experience an increase in utility.

$$\frac{\hat{x}}{q_2} = \left(\frac{\beta(23\beta^2 - 38\beta\lambda + 20\lambda^2)}{(\beta - \lambda)(23\beta^2 + 4\beta\lambda - \lambda^2)} \right) \qquad \textbf{(A4)}$$

This ratio is positive as long as the condition for product differentiation ($\lambda<0.5\beta$) holds. Thus some positive proportion of HMO 2 members are hurt by an increase in relative practice norms.